



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :

G06K 9/20 // G06F 17/24

A1

(11) International Publication Number:

WO 98/47098

(43) International Publication Date:

22 October 1998 (22.10.98)

(21) International Application Number: PCT/SE98/00602

(22) International Filing Date: 1 April 1998 (01.04.98)

(30) Priority Data:

9701183-7

1 April 1997 (01.04.97)

SE

(71) Applicant (for all designated States except US): READSOFT AB [SE/SE]; Garnisionsgatan 25A, S-254 66 Helsingborg (SE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): ANDERSSON, Jan [SE/SE]; Banckagatan 17, S-260 40 Viken (SE).

(74) Agents: ONN, Thorsten et al.; AB Stockholms Patentbyrå, Zacco & Bruhn (publ), P.O. Box 23101, S-104 35 Stockholm (SE).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

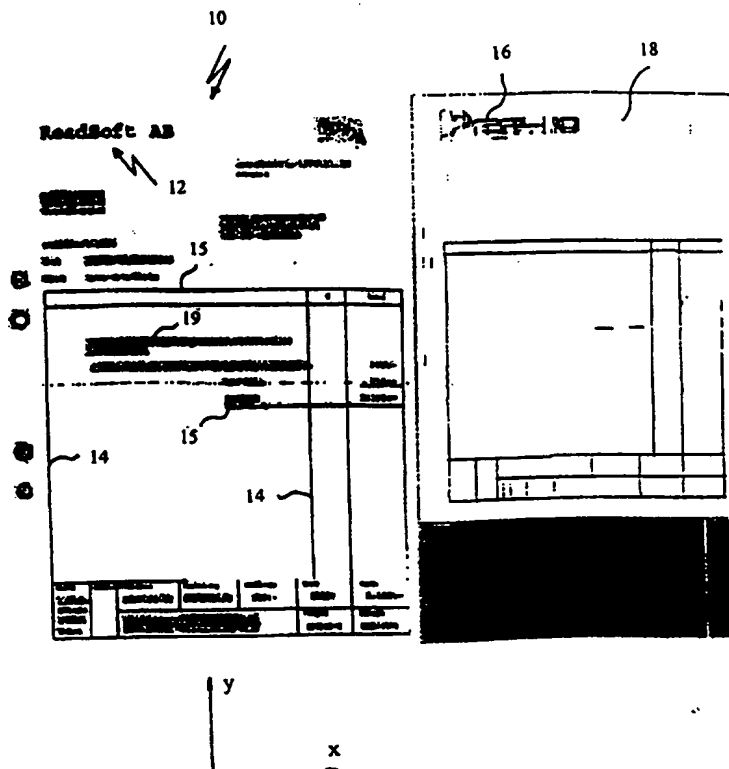
With international search report.

In English translation (filed in Swedish).

(54) Title: METHOD AND ARRANGEMENT FOR AUTOMATIC DATA ACQUISITION OF FORMS

(57) Abstract

The invention relates to a method and an arrangement for the automatic data acquisition of paper-based information, whereby an unknown form (10) that is scanned into a computer is transformed into line elements (14, 15) for quicker identification. The identification of known forms (10) is also simplified by means of the line elements (14, 15).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Title

Method and arrangement for automatic data acquisition of forms

Technical field

5 The present invention refers to a method and arrangement for the automatic data acquisition, by means of a means for the same, of forms whose design and information is not known in advance, by input into the said means together with storage of patterns of the same.

The prior art

It is a problem for companies, organisations and others to make good use of the information found in different types of paper forms, documents, etc.

10 With new, modern technology, these items can be scanned with a scanner and entered into a database via commercially available software programs. However, sorting, identification and other checking routines must to a large extent still be performed manually via the computer's display or screen.

15 For example, to store an invoice from one and the same company as one specifically designed document with a logotype and other visual elements, it must be revised so that its format is adapted to one that can be accepted by the software and then stored in a database. This and other procedures must be repeated each time an invoice with a new design is scanned with the software.

20 To identify an invoice from a company that is already registered, the whole invoice is often searched, which is time consuming. Certain software programs can have search routines that restrict the extent of this searching. It is, however, difficult to safeguard against blurred or hand-written lines of text, etc.

25 A need therefore exists for all who handle invoices and other forms to quickly be able to identify these and/or quickly be able to enter and store new patterns in their invoicing system.

Patent US-A-4 933 979 describes traditional data acquisition from forms and requires pre-defined templates/patterns with no self-learning (adaptive) ability.

30 Patent US-A-5 140 650 mentions data acquisition from forms with what is known as "Form out" technology to cover-over the original document and only retain the parts that are "filled-in". This data acquisition is often combined with data acquisition according to US-A-4 933 979. The patent does not have any adaptive function for data acquisition of unknown forms.

Another patent, US-A-5 293 429, concerns the classification of documents with the help of lines on the documents and does not directly concern data acquisition or any adaptive function for this. US-A-5 293 429 does not ensure the identification of lines with object areas (areas with text) and a "RCG-value" (ReCoGnition, a number that uniquely identifies a document).

None of the said patents generates a form map for a form unknown to the system according to the patent and stores the map in real time in a form database for recognition at the next opportunity for identification. For the inventions according to these patents, the unknown form must therefore be stored later by other means.

Summary of the invention

One of the objectives of the present invention is to solve the problems named above as well as others during what is known as automatic data acquisition (interpretation) in connection with the handling of paper-based information.

The present invention concerns a system (method and arrangement) for the automatic data acquisition of forms where the system has no prior knowledge of what the form looks like or where on the form the information is to be found. In this way, templates of forms do not have to be defined in advance, but are instead registered as they are submitted to the system, i.e., in real time.

To accomplish the above objectives, the present invention specifies a method and arrangement for the automatic data acquisition, by means of a means for the same, of forms whose design and information is not known in advance, by input into the said means together with storage of patterns of the same. The method is adaptive, by which it includes learning and registering of forms as patterns without filled-in text, and by it also including the following steps for accomplishing the adaptive registration:

generation of a form map based on the previously unknown form's design for identifying information contained on the form;

searching and comparing the form map with stored, registered maps in a means for storing form maps;

storing generated form maps in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

indication of agreement according to the limits for agreement when agreement is found; and

continued data acquisition for identifying of the information content of the form.

According to one embodiment of the present invention, the form map can consist of an object area list with objects contained in the form whereby the object comprises colours and/or wholly or partly of text.

5 In an alternative embodiment, the form map constitutes a line map comprising objects in the form of coloured lines from the form.

Horizontal lines in the line map are used to produce a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

10 Vertical lines in the line map are used to produce a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

At least one line element that is included in a segment is marked in the equivalent key position, and segments that lack line elements remain unmarked in the equivalent key position.

15 A horizontal key and/or a vertical key constitute a line key in the line map, whereby during the said searching, the line key generated is compared with line keys stored in the means for verifying agreement.

The line keys are sorted in the storage means according to the number of markings.

20 The object's horizontal position in the object area list is used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

25 The object's vertical position in the object area list is used to produce a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

At least one object that is included in a segment is marked in the equivalent key position, and segments that lack objects remain unmarked in the equivalent key position.

30 A horizontal key and/or a vertical key constitute an object key in the object area list, whereby during the said searching, the object key generated is compared with object keys stored in the means for verifying agreement.

The object keys are preferably sorted in the storage means according to the number of markings.

Searching results in a pre-defined number of requested probable candidates for the currently searched form.

5 If needed, an operator can support manually the whole or parts of the adaptive registration or identification of the new form or registered forms respectively if several alternative candidates are found as probabilities according to a factor of merit.

Finally, the identity of the form is confirmed by the data acquisition of a RCG-value.

10 Furthermore, the present invention specifies a arrangement for performing the above method.

The arrangement carries out automated data acquisition, by means of a means for the same, of forms whose design and information is not known in advance, by input into the said arrangement together with storage of patterns of the same. It learns adaptively and registers the design of the form, and includes a computer with the following means for carrying out the adaptive registration:

means for generation of a form map based on the previously unknown form's design for identifying information contained on the form;

20 means for searching and comparing the form map with stored, registered maps in a means for storing form maps;

means for storing generated form maps in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

means for indicating agreement according to the limits for agreement when agreement is found; and

25 means for identification and continued data acquisition of the information content of the form.

In addition, the arrangement can include or constitute that specified according to the above method of the present invention, which is further illustrated in the accompanying non-independent claims for the arrangement.

30 **Brief description of the drawings**

Further reference to the enclosed figures and associated text will give a clearer understanding of the present invention.

Fig. 1 shows schematically how a line pattern is accomplished from a scanned-in invoice.

Fig. 2 shows schematically a flow-path for scanning, identifying, interpreting and validating a form according to the present invention.

Detailed description of preferred embodiments

In the continued description of the present invention, the forms are presented as invoices. However, the invention is not limited to invoice forms but also covers general documents containing text, figures, etc. as forms. Invoices are used here as an example of a form to exemplify the invention.

Fig. 1 illustrates schematically one part of invoice 10 that is scanned in a computer and that is shown on the display. As is evident from invoice 10, it is unclear or blurred after the scanning or input.

Invoice 10 consists partly of a logotype 12 and the vertical 14 and horizontal line 15 elements.

Note that even the logotype contains long black or varying degrees of shaded coloured line elements 16 that have been partly registered in a line map 18 according to the present invention, and that give an idea of what the original logotype 12 looked like, which simplifies identification when the invoice is an object to be identified as being as registered in a form map database. Coloured lines also include grey scales of black.

The form map that in this case constitutes line map 18 has been filtered from other objects 19, such as whole or parts of text objects or coloured objects, plus even the said line elements that include colour, which cannot be reproduced here, but that can be included as many coloured fields on a form 10.

An invoice 10 that is prepared according to the present invention, hereafter designated EH (Eyes & Hands), must be identified at an early stage. For successful identification, EH must have on a previous occasion, learned what the current invoice 10 looks like, which in simple terms means that information about the invoice is available in the form database in EH.

By necessity, the identification must be quick and be able to be made in a database that holds a very large number of invoice identities 18. It is not uncommon for databases to contain more than 10,000 identities 18.

The method and arrangement that EH uses does not require that an invoice is always put through a scanner in exactly the same way, i.e. the information on the invoice can vary somewhat in the x and y axes within a pre-determined measurement or threshold value. Fig. 1 shows a schematic cartographic system of co-ordinates.

5 In the present invention, (EH) comprises that in one embodiment, EH searches for all vertical 14 and horizontal line 15 elements of a pre-determined length on the invoice. Lines 14, 15 do not need to be free-standing and isolated, but can, for example, be part of a larger logotype text 12, such as ReadSoft AB in Fig. 1. The logotype 12 is represented as the line element 16 in line map 18.

10 The horizontal lines 15 and the vertical lines 14 constitute the basis for generating a horizontal key (h-key) and a vertical key (v-key) respectively according to the following:

* The invoice is divided into a large number of horizontal segments along the y-axis (not shown). Each segment is equivalent to one position in the h-key. If a certain segment includes one or more line elements 15, a mark or tag is placed in the equivalent key position. If not, an empty space, an inverted mark or anything else that differentiates itself from a mark is used instead.

* A v-key for the vertical line elements 14 is generated in a similar manner along the x-axis.

20 * The h and v-keys are given designations and together constitute a line key. Following this, a search is performed, which means that the current line key is compared with line keys for known invoices 10 that exist in EH's database. This comparison takes into account that individual lines or line elements 14, 15 can vary somewhat in position, plus that the total pattern of lines can be displaced somewhat according to suitable pre-determined values in the x and y directions, horizontally and vertically respectively.

25 * The line keys in the database are sorted according to the number of markings (tags), which are used to make the searching effective.

* The search results in a pre-determined number of probable candidates for the identity of the current invoice 10. All candidates are associated with a factor of merit or a probability that they are the current invoice 10.

30 * The identity of the invoice is finally confirmed by carrying out an interpretation of that known as the RCG-value (RCG- ReCoGnition). The RCG-value is a value at a given position that is unique for a certain invoice/supplier or other form. Examples of such

values are bank giro numbers, post-office giro numbers, invoice numbers, total amounts, etc.

The said segments can, for example, form checked patterns that are fine-screened to varying extents according to the relative need for rapid searching.

5 The line keys can even be implemented on objects formed wholly or partly of text and colours. These are assigned line keys from an object area list that includes x and y-keys for the object. The object area list can, for example, consist of positions for certain selected objects. The principles for line maps stated above are even appropriate for objects other than line elements to accomplish identification of forms.

10 If the line keys are not found in the database, this indicates that the invoice is not known, which results in the new line keys being stored in the database that, in this way, is updated in real time.

 If necessary, the operator can, via his computer, manually support the whole or part of the adaptive registration and/or identification of a new form or registered form respectively if
15 several alternative candidates are presented as probable according to the factor of merit.

 In addition, the present invention includes an arrangement for performing the method according to the above.

 The arrangement performs the automatic data acquisition, by means of a means for the same, of forms whose design and information content is not known in advance, by input into
20 the said means together with storage of patterns of the same. It registers in an adaptive manner and learns the design of forms, and includes a computer with the following means for accomplishing the adaptive registration:

 means for generating a form map based on the previously unknown form's design for identifying information contained on the form;

25 means for searching and comparing the form map with stored, recognised maps in a means for storing form maps;

 means for storage of generated form maps in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

 means for indicating agreement according to the limits for agreement when
30 agreement is found; and

 means for identification and continued data acquisition of the information content of the form.

The said means are preferably controlled by computer hardware and software, such as, for example:

A scanner for acquisition of data.

An electronic storage medium (hard disk, CD-ROM, etc.) for the means to store information

Signs, icons, signal generators, etc. for indicating purposes.

Filters and comparitors so that the means can search and compare, as well as filters and registers for identification.

On the whole, the means used in the present invention are well known to a skilled person in the technical field, but the way in which they are co-ordinated to achieve the object of the invention is, however, innovative.

In one embodiment of the present invention with reference to Fig. 2, a schematic flow-path is illustrated to show the scanning, identification, interpretation, and validation of a form according to the present invention.

Fig. 2 is divided by dotted lines into partial areas to clarify the different steps in a method according to the invention, whereby the steps constitute the scanning of the form 200, identification of the form 210, interpretation of the form 220, plus validating the form 230.

The form is scanned 200 into EH, and identification 210 follows. Identification consists of generating a line map 212, or alternatively an object area list, whereby an line key is generated. Following this, form 10 is compared 214 with known keys in the form map database, whereby a conformation of identification is obtained via the RCG-value. The next step includes deciding whether the identification was successful 216 according to the conditions "Yes" or "No". If the decision results in "No", a conditional investigation is made to see if there are more candidates in the form of line keys 218. In the answer here is "Yes", a loop in the form of 214, 216 and 218 is performed until a successful identification is finally made, or until no further line key candidates are presented 218.

In the case of a successful identification, interpretation 220 of the form then begins by interpreting with the help of the current form map 222, after which validation 230 or evaluation 232 of the fields of the form 10 takes place. As an option, the operator can assist with selection if several alternative fields are found 234.

If the identification 210 is unsuccessful, and no further line keys are presented 218, interpretation 220 is performed in that self-learning with a form definition 224 is accomplished.

The form definition consists of a template or a set of rules that describes the common elements of a specific collection of forms, for example, Swedish invoices. Following this, the RCG-value is interpreted 226 and a decision is made 228 whether the current RCG-value can be found in the form database. If the answer is "Yes", a re-interpretation begins 229, followed by a continued
5 interpretation 222 that leads to validation 232.

If, on the other hand, the answer is "No", validation commences 230, 236, after which the form is saved in the form map database with the line key 238. Prior to steps 236, 238, the operator can, if several field alternatives are found, assist with the self-learning process.

The embodiments of the present invention described above are possible
10 embodiments, but are not intended to limit the invention to such, as further embodiments will be evident to a skilled person in the technical area via the drafts of the enclosed claims.

Claims

1. Method for the automatic data acquisition (200), by means of a means for the same, of forms (10) whose design and information content (19) is not known in advance, by input into the said means, together with storage of patterns of the same, characterised in that the method is adaptive (224) and includes self-learning and registration of the design of forms, whereby it includes the following steps to accomplish the adaptive registration (238):

generation of a form map (18) based on the previously unknown form's (10) design for identifying (210) information contained on the form;

searching and comparing (210, 220) the form map (18) with stored, registered maps in a means for storing form maps;

storage (238) of generated form maps (18) in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

indication of agreement according to the limits for agreement when agreement is found; and

continued data acquisition (232) for identifying the information content of the form.

2. Method according to claim 1, characterised in that the form map (18) consists of an object area list with objects (19) contained in the form.

3. Method according to claim 2, characterised in that the object (19) comprises colours and/or wholly or partly of text.

4. Method according to claim 1, characterised in that the form map (18) constitutes a line map comprising line elements (14,15) from the form (10).

5. Method according to claim 4, characterised in that the horizontal lines (15) in the line map are used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of coordinates, whereby each segment is equivalent to a position in the horizontal key.

6. Method according to claims 4 and 5, characterised in that the vertical lines (14) in the line map are used to generate a vertical key by dividing the form (10) into a pre-determined number of vertical segments along the x-axis in a cartographic system of coordinates, whereby each segment is equivalent to a position in the vertical key.

7. Method according to claims 5 and 6, characterised in that at least one line element (14, 15) that is included in a segment is marked in the equivalent key position, and that segments that lack line elements remain unmarked in the equivalent key position.

8. Method according to claims 4-7, characterised in that the horizontal key and/or a vertical key constitute a line key in the line map (18), whereby, during the said searching, the line key generated is compared (214) with line keys stored in the storage means for verifying agreement.

5 9. Method according to claim 8, characterised in that the line keys are sorted in the storage means according to the number of markings.

10 10. Method according to claims 1 and 2, characterised in that the object's horizontal position in the object area list is used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

11. Method according to claims 1 and 2, characterised in that the object's vertical position in the object area list is used to generate a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

15 12. Method according to claims 10 and 11, characterised in that at least one object that is included in a segment is marked in the equivalent key position, and that segments that lack objects remain unmarked in the equivalent key position

20 13. Method according to claims 10-12, characterised in that a horizontal key and/or a vertical key constitute an object key in the object area list, whereby, during the said searching, the object key generated is compared with object keys stored in the storage means for verifying agreement.

14. Method according to claim 13, characterised in that the object keys are sorted in the storage means according to the number of markings.

25 15. Method according to claims 1-14, characterised in that the searching (210, 218) results in a pre-defined number of requested probable candidates for the currently searched form.

30 16. Method according to claim 15, characterised in that an operator can support manually (234, 239) the whole or parts of the adaptive registration or identification of the new form or registered forms respectively if several alternative candidates are found as probabilities according to a factor of merit.

17. Method according to claims 1-16, characterised in that the identity of the form is confirmed by the data acquisition of a RCG-value (214).

18. Arrangement for the automatic data acquisition, by means of a means for the same, of forms (10) whose design and information content (19) is not known in advance, by input into the said means together with storage of patterns of the same, characterised in that it learns adaptively and registers the design of forms (10), and includes a computer with the
5 following means for carrying out the adaptive registration (238):

means for generating a form map (18) based on the previously unknown form's (10) design for identifying information contained on the form;

means for searching and comparing the form map with stored, recognised maps in a means for storing form maps;

10 means for storage of generated form maps (18) in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

means for indicating agreement according to the limits for agreement when agreement is found; and

15 means for identification and continued data acquisition of the information content of the form.

19. Arrangement according to claim 18 characterised in that the form map (18) consists of an object area list with objects (19) contained in the form (10).

20. Arrangement according to claim 19, characterised in that the object (19) comprises colours and/or wholly or partly of text.

20 21. Arrangement according to claim 18, characterised in that the form map (18) constitutes a line map comprising line elements (14, 15) from the form.

22. Arrangement according to claim 21, characterised in that the horizontal lines (15) in the line map are used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.
25

23. Arrangement according to claims 21 and 22, characterised in that the vertical lines (14) in the line map are used to generate a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

30 24. Arrangement according to claims 22 and 23, characterised in that at least one line element (14, 15) that is included in a segment is marked in the equivalent key

position, and that segments that lack line elements remain unmarked in the equivalent key position.

25. Arrangement according to claims 22-24, characterised in that the horizontal key and/or a vertical key constitute a line key in the line map, whereby, during the said searching, the line key generated is compared with line keys stored in the storage means for verifying agreement.

26. Arrangement according to claim 25, characterised in that the line keys are sorted in the storage means according to the number of markings.

27. Arrangement according to claims 18 and 19, characterised in that the object's (19) horizontal position in the object area list is used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

28. Arrangement according to claims 18 and 19, characterised in that the object's (19) vertical position in the object area list is used to generate a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

29. Arrangement according to claims 27 and 28, characterised in that at least one object that is included in a segment is marked in the equivalent key position, and that segments that lack objects remain unmarked in the equivalent key position.

30. Arrangement according to claims 27-29, characterised in that a horizontal key and/or a vertical key constitute an object key in the object area list, whereby, during the said searching, the object key generated is compared with object keys stored in the storage means for verifying agreement.

31. Arrangement according to claim 30, characterised in that the object keys are sorted in the storage means according to the number of markings.

32. Arrangement according to claims 18-31, characterised in that the searching results in a pre-defined number of requested probable candidates for the currently searched form.

33. Arrangement according to claim 32, characterised in that an operator can support manually the whole or parts of the adaptive registration or identification of the new

form or registered forms respectively if several alternative candidates are found as probabilities according to a factor of merit.

34. Arrangement according to claims 18-33, characterised in that the identity of the form is confirmed by the data acquisition of a RCG-value.

5 -----

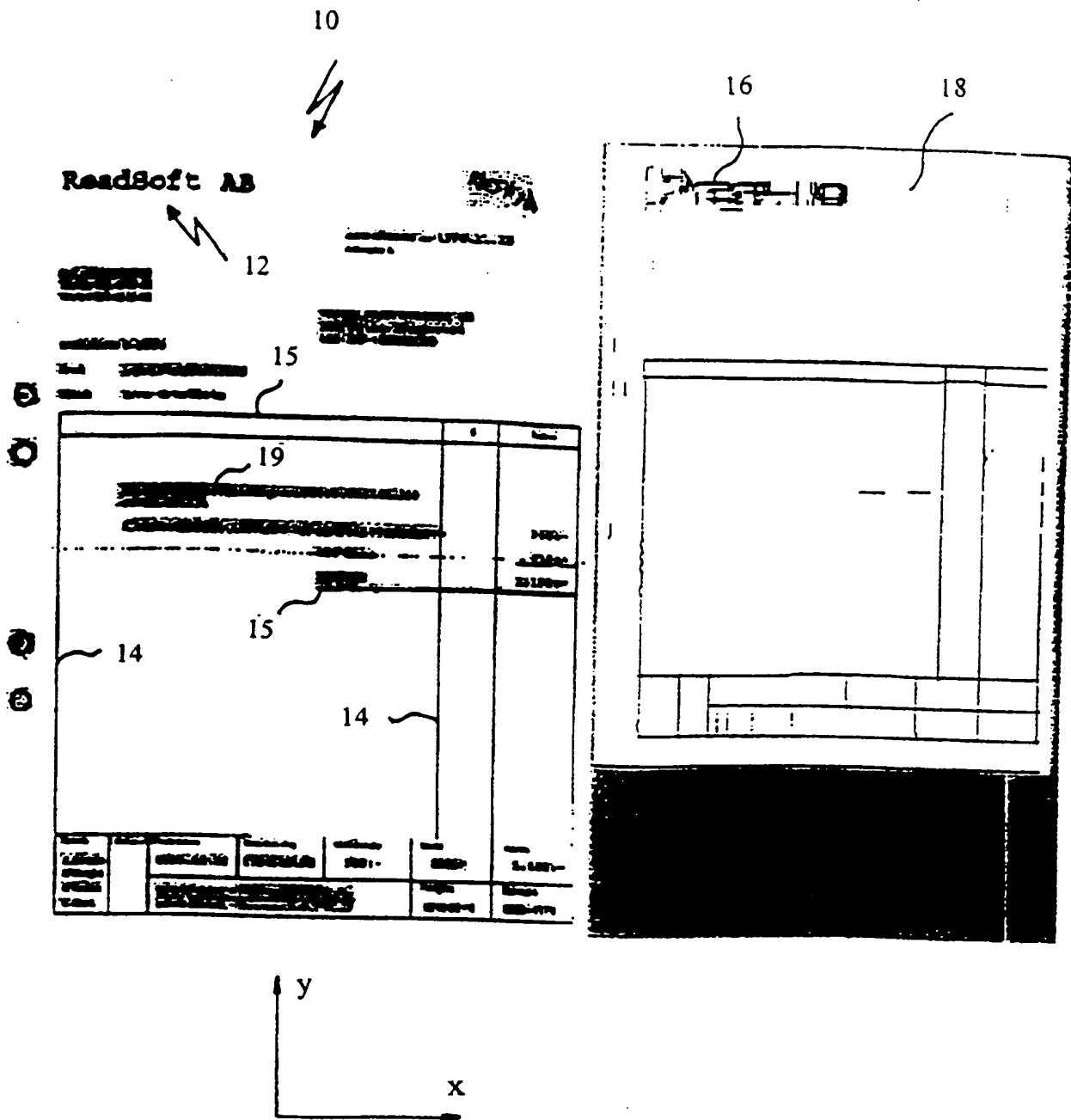


Fig. 1

2/2

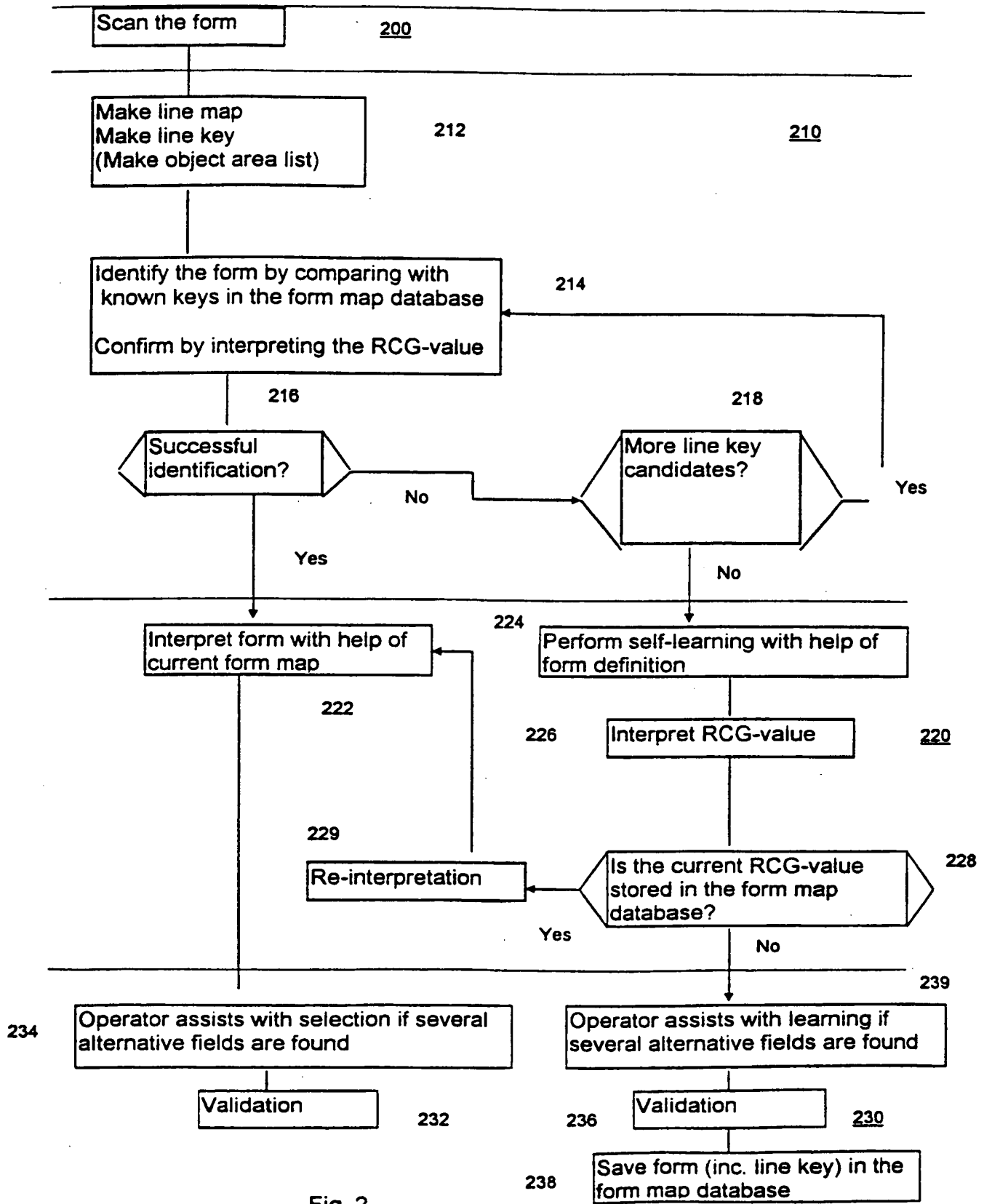


Fig. 2

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No. PCT/SE 98/00602

International Filing Date

01-04-1998

**The Swedish Patent Office
PCT International Application**

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference
(if desired) (12 characters maximum) P33354PC00

Box No. I TITLE OF INVENTION Method and arrangement for automatic data acquisition of forms	
Box No. II APPLICANT <div style="display: flex; justify-content: space-between;"> <div style="width: 60%;"> <p><small>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)</small></p> <p>ReadSoft AB Garnisionsgatan 25A S-254 66 HELSINGBORG Sweden</p> </div> <div style="width: 35%;"> <p><input type="checkbox"/> This person is also inventor.</p> <p>Telephone No.</p> <p>Facsimile No.</p> <p>Teleprinter No.</p> </div> </div> <div style="display: flex; justify-content: space-between; margin-top: 10px;"> <p><small>State (i.e. country) of nationality:</small> SE</p> <p><small>State (i.e. country) of residence:</small> SE</p> </div> <p><small>This person is applicant for the purposes of:</small> <input type="checkbox"/> all designated States <input checked="" type="checkbox"/> all designated States except the United States of America <input type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box</p>	
Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S) <div style="display: flex; justify-content: space-between;"> <div style="width: 60%;"> <p><small>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)</small></p> <p>ANDERSSON; Jan Banckagatan 17 S-260 40 VIKEN Sweden</p> </div> <div style="width: 35%;"> <p><small>This person is:</small></p> <p><input type="checkbox"/> applicant only</p> <p><input checked="" type="checkbox"/> applicant and inventor</p> <p><input type="checkbox"/> inventor only (If this check-box is marked, do not fill in below.)</p> </div> </div> <div style="display: flex; justify-content: space-between; margin-top: 10px;"> <p><small>State (i.e. country) of nationality:</small> SE</p> <p><small>State (i.e. country) of residence:</small> SE</p> </div> <p><small>This person is applicant for the purposes of:</small> <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box</p> <p><input type="checkbox"/> Further applicants and/or (further) inventors are indicated on a continuation sheet.</p>	
Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE <p><small>The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:</small> <input checked="" type="checkbox"/> agent <input type="checkbox"/> common representative</p> <div style="display: flex; justify-content: space-between;"> <div style="width: 60%;"> <p><small>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)</small></p> <p>AB STOCKHOLMS PATENTBYRÅ, Zacco & Bruhn (publ) ONN, Thorsten; AGVALD-GLAS, Gunilla; BERNHULT, Lennart; HANSSON, Sven; PETRÉ, Urban; WESTERLUND, Örjan</p> <p>Box 23101, S-104 35 STOCKHOLM, Sweden</p> </div> <div style="width: 35%;"> <p>Telephone No. 08-7299500</p> <p>Facsimile No. 08-318315</p> <p>Teleprinter No.</p> </div> </div> <p><input type="checkbox"/> Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.</p>	

Box No.V DESIGNATION OF STATE

01-04-1998

The following designations are hereby made under Rule 4.9(a) (mark the applicable check-boxes; at least one must be marked):

Regional Patent

- ☒ AP ARIPO Patent: GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SZ Swaziland, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☒ EA Eurasian Patent: AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☒ EP European Patent: AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☒ OA OAPI Patent: BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT (if other kind of protection or treatment desired, specify on dotted line)

National Patent (if other kind of protection or treatment desired, specify on dotted line):

- | | |
|--|--|
| <input checked="" type="checkbox"/> AL Albania | <input checked="" type="checkbox"/> LT Lithuania |
| <input checked="" type="checkbox"/> AM Armenia | <input checked="" type="checkbox"/> LU Luxembourg |
| <input checked="" type="checkbox"/> AT Austria | <input checked="" type="checkbox"/> LV Latvia |
| <input checked="" type="checkbox"/> AU Australia | <input checked="" type="checkbox"/> MD Republic of Moldova |
| <input checked="" type="checkbox"/> AZ Azerbaijan | <input checked="" type="checkbox"/> MG Madagascar |
| <input checked="" type="checkbox"/> BA Bosnia and Herzegovina | <input checked="" type="checkbox"/> MK The former Yugoslav Republic of Macedonia |
| <input checked="" type="checkbox"/> BB Barbados | |
| <input checked="" type="checkbox"/> BG Bulgaria | <input checked="" type="checkbox"/> MN Mongolia |
| <input checked="" type="checkbox"/> BR Brazil | <input checked="" type="checkbox"/> MW Malawi |
| <input checked="" type="checkbox"/> BY Belarus | <input checked="" type="checkbox"/> MX Mexico |
| <input checked="" type="checkbox"/> CA Canada | <input checked="" type="checkbox"/> NO Norway |
| <input checked="" type="checkbox"/> CH and LI Switzerland and Liechtenstein | <input checked="" type="checkbox"/> NZ New Zealand |
| <input checked="" type="checkbox"/> CN China | <input checked="" type="checkbox"/> PL Poland |
| <input checked="" type="checkbox"/> CU Cuba | <input checked="" type="checkbox"/> PT Portugal |
| <input checked="" type="checkbox"/> CZ Czech Republic | <input checked="" type="checkbox"/> RO Romania |
| <input checked="" type="checkbox"/> DE Germany | <input checked="" type="checkbox"/> RU Russian Federation |
| <input checked="" type="checkbox"/> DK Denmark | <input checked="" type="checkbox"/> SD Sudan |
| <input checked="" type="checkbox"/> EE Estonia | <input checked="" type="checkbox"/> SE Sweden |
| <input checked="" type="checkbox"/> ES Spain | <input checked="" type="checkbox"/> SG Singapore |
| <input checked="" type="checkbox"/> FI Finland | <input checked="" type="checkbox"/> SI Slovenia |
| <input checked="" type="checkbox"/> GB United Kingdom | <input checked="" type="checkbox"/> SK Slovakia |
| <input checked="" type="checkbox"/> GE Georgia | <input checked="" type="checkbox"/> SL Sierra Leone |
| <input checked="" type="checkbox"/> GH Ghana | <input checked="" type="checkbox"/> TJ Tajikistan |
| <input checked="" type="checkbox"/> GM Gambia | <input checked="" type="checkbox"/> TM Turkmenistan |
| <input checked="" type="checkbox"/> GW Guinea-Bissau | <input checked="" type="checkbox"/> TR Turkey |
| <input checked="" type="checkbox"/> HU Hungary | <input checked="" type="checkbox"/> TT Trinidad and Tobago |
| <input checked="" type="checkbox"/> ID Indonesia | <input checked="" type="checkbox"/> UA Ukraine |
| <input checked="" type="checkbox"/> IL Israel | <input checked="" type="checkbox"/> UG Uganda |
| <input checked="" type="checkbox"/> IS Iceland | <input checked="" type="checkbox"/> US United States of America |
| <input checked="" type="checkbox"/> JP Japan | |
| <input checked="" type="checkbox"/> KE Kenya | <input checked="" type="checkbox"/> UZ Uzbekistan |
| <input checked="" type="checkbox"/> KG Kyrgyzstan | <input checked="" type="checkbox"/> VN Viet Nam |
| <input checked="" type="checkbox"/> KP Democratic People's Republic of Korea | <input checked="" type="checkbox"/> YU Yugoslavia |
| | <input checked="" type="checkbox"/> ZW Zimbabwe |
| <input checked="" type="checkbox"/> KR Republic of Korea | |
| <input checked="" type="checkbox"/> KZ Kazakhstan | |
| <input checked="" type="checkbox"/> LC Saint Lucia | |
| <input checked="" type="checkbox"/> LK Sri Lanka | |
| <input checked="" type="checkbox"/> LR Liberia | |
| <input checked="" type="checkbox"/> LS Lesotho | |

Check-boxes reserved for designating States (for the purposes of a national patent) which have become party to the PCT after issuance of this sheet:

- ☐
- ☐
- ☐

In addition to the designations made above, the applicant also makes under Rule 4.9(b) all designations which would be permitted under the PCT except the designation(s) of

The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. (Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)

Box No. VI PRIORITY CLAIM

Further priority claims are indicated in the Supplemental Box ☐

The priority of the following earlier application(s) is hereby claimed:

Country (in which, or for which, the application was filed)	Filing Date (day/month/year)	Application No.	Office of filing (only for regional or international application)
item (1) SWEDEN	(01.04.1997) 1 APRIL 1997	9701183-7	
item (2)			
item (3)			

Mark the following check-box if the certified copy of the earlier application is to be issued by the Office which for the purposes of the present international application is the receiving Office (a fee may be required):

☒ The receiving Office is hereby requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) identified above as item(s): (9701183-7) (1)

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) (If two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used): ISA / SE

Earlier search Fill in where a search (international, international-type or other) by the International Searching Authority has already been carried out or requested and the Authority is now requested to base the international search, to the extent possible, on the results of that earlier search. Identify such search or request either by reference to the relevant application (or the translation thereof) or by reference to the search request.

Country (or regional Office):

Date (day/month/year):

Number:

Box No. VIII CHECK LIST

This international application contains the following number of sheets:

1. request : 3 sheets
 2. description : 9 sheets
 3. claims : 5 sheets
 4. abstract : 1 sheet
 5. drawings : 2 sheets
 Total : 20 sheets

This international application is accompanied by the item(s) marked below:

1. ☐ separate signed power of attorney
 2. ☐ copy of general power of attorney
 3. ☐ statement explaining lack of signature
 4. ☐ priority document(s) identified in Box No. VI as item(s):
 5. ☒ fee calculation sheet
 6. ☐ separate indications concerning deposited microorganisms
 7. ☐ nucleotide and/or amino acid sequence listing (diskette)
 8. ☐ other (specify):

Figure No. 1 of the drawings (if any) should accompany the abstract when it is published.

Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

Sven A. Hansson

Sven A. Hansson

Representative of the applicant

For receiving Office use only

1. Date of actual receipt of the purported international application:

01-04-1998

3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:

4. Date of timely receipt of the required corrections under PCT Article 11(2):

5. International Searching Authority specified by the applicant:

ISA / SE

6. ☐ Transmittal of search copy delayed until search fee is paid

2. Drawings:

☒ received:☐ not received:

Date of receipt of the record copy by the International Bureau:

27 APRIL 1998

27.04.98

Förfarande och anordning för automatisk datafångst hos formulär

Tekniskt område

- 5 Föreliggande uppfinning hänför sig till ett förfarande och en anordning för automatisk datafångst, via organ för detsamma, av formulär med på förhand okänd utformning och informationsinnehåll, vid inmatning i nämnda organ, samt lagring av förlagor av densamma.

Teknikens standpunkt

- 10 Det innebär problem för företag, organisationer och andra, att ta vara på information som förekommer på olika typer av formulär, blanketter, dokument etc.

- Med ny och modern teknik avsöks sådana med en skanner och skannas in i en databas via på marknaden förekommande datorprogram. Sortering, identifiering och andra kontrollrutiner måste dock i hög grad skötas manuellt interaktivt via datorns display eller
15 skärm.

- För att t ex lagra en faktura, som kan förekomma som ett, för ett och samma företag, specifikt utformat dokument med logotyp och andra förtecken, måste den omarbetas så att dess format anpassas till vad programvaran accepterar och sedan lagras i en databas. Varje gång en faktura med ny utformning skannas in i programvaran måste
20 dessa och andra procedurer upprepas.

För att identifiera en faktura från ett företag som redan finns registrerat söks ofta på hela fakturans utformning, vilket är tidskrävande. Vissa programvaror kan ha sökrutiner som begränsar dess omfattning. Det är dock svårt att gardera mot blurriga textsträngar, handskrivna sådana etc.

- 25 Det existerar därför ett behov, hos alla som hanterar fakturor eller andra blanketter, att snabbt kunna identifiera dessa och/eller att snabbt kunna införa och lagra nya förlagor i sitt faktureringsystem.

Patentet US-A-4 933 979 beskriver traditionell datafångst från formulär och kräver fördefinierade mallar/mönster utan någon självlärande (adaptiv) förmåga.

- 30 I patentet US-A-5 140 650 anges datafångst från formulär med s k "Form out" teknik för att maska bort ursprungsdokumentet och endast behålla de "ifyllda delarna", vilken datafångst ofta kombineras med datafångsten enligt US-A-4 933 979. Patentet

uppvisar inte någon adaptiv funktion för datafångst av okända formulär.

Ett annat patent US-A-5 293 429 avser klassificering av dokument med hjälp av linjer på dokumenten och berör inte direkt datafångst eller någon adaptiv funktion för detta. US-A-5 293 429 säkerställer inte identifiering av linjer med objektområden

- 5 (områden med text) och ett "RCG-värde" (ReCoGnition, nummer som unikt identifierar ett dokument).

- 10 Inget av här nämnda patentedokument alstrar en formulärkarta av ett för systemen enligt patenten okänt formulär och lagrar kartan i realtid så att den ingår i en formulärdata-tabas för igenkännande vid nästa tillfälle för identifiering. För uppfinningarna enligt patenten måste således dokument som är okända lagras i efterhand med separata metoder.

Sammanfattning av uppfinningen

Ett av ändamålen med föreliggande uppfinning är att lösa ovan nämnda och andra problem vid s k automatisk datafångst (tolkning) i samband med hantering av pappersburen information.

- 15 Föreliggande uppfinning avser ett system (förfarande och en anordning) för automatisk datafångst på formulär där systemet inte från början känner till hur ett formulär ser ut och var på formuläret information återfinnes. På så vis behöver inte formulärmallar definieras i förväg, utan dessa registreras allt eftersom de inkommer i systemet, d v s i realtid.

- 20 För att åstadkomma ovanstående anger föreliggande uppfinning ett förfarande för automatisk datafångst, via organ för detsamma, av formulär med på förhand okänd utformning och informationsinnehåll, vid inmatning i nämnda organ, samt lagring av förlagor av densamma. Förfarandet är adaptivt, varvid det innefattar inläring och registrering av utformningen av formulär som förlagor utan ifylld text, och varvid det
- 25 vidare innefattar följande steg för att åstadkomma den adaptiva registreringen:

alstring av en formulärkarta utifrån det på förhand okända formulärets utformning för identifiering av formulär innehållande information;

sökning med jämförelse av formulärkartan med lagrade inlärd kartor i ett organ för lagring av formulärkartor;

- 30 lagring av alstrad formulärkarta i organet för lagring då den ej överensstämmer med en lagrad karta enligt förutbestämda gränser för överensstämmelse;

indikering av överensstämmelse enligt gränserna för överensstämmelse då

överensstämmelse föreligger; och

fortsatt datafångst för identifiering av formulärets informationsinnehåll.

Formulärkartan kan i en utföringsform av föreliggande uppfinning bestå av en objektområdeslista med objekt innefattade i formulär, varvid objekten består av färger
5 och/eller hel eller del av text.

I en alternativ utföringsform utgör formulärkartan en linjekarta innefattande objekt i form av färgade linjeelement ur formuläret.

Horisontella linjer i linjekartan används för alstring av en horisontell nyckel genom att formuläret indelas i ett förutbestämt antal horisontella segment längs y-axeln i
10 ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den horisontella nyckeln.

Vertikala linjer i linjekartan används för alstring av en vertikal nyckel genom att formuläret indelas i ett förutbestämt antal vertikala segment längs x-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den vertikala nyckeln.

15 Minst ett linjeelement som innefattas i ett segment märks i motsvarande nyckelposition och segment vilka saknar linjeelement förblir omärkta i motsvarande nyckelposition.

En horisontell nyckel och/eller en vertikal nyckel utgör en linjenyckel i linjekartan, varvid, vid nämnda sökning, alstrad linjenyckel jämförs med linjenycklar som
20 finns lagrade i organet för lagring för kontroll av överensstämmelse.

Linjenycklarna är sorterade efter antalet märkningar i organet för lagring.

Objekts horisontella position i objektområdeslistan används för alstring av en horisontell nyckel genom att formuläret indelas i ett förutbestämt antal horisontella segment längs y-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en
25 position i den horisontella nyckeln.

Objekts vertikala position i objektområdeslistan används för alstring av en vertikal nyckel genom att formuläret indelas i ett förutbestämt antal vertikala segment längs x-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den vertikala nyckeln.

30 Minst ett objekt som innefattas i ett segment märks i motsvarande nyckelposition och segment vilka saknar objekt förblir omärkta i motsvarande nyckelposition.

En horisontell nyckel och/eller en vertikal nyckel utgör en objektnyckel i objekt-

områdeslistan, varvid, vid nämnda sökning, alstrad objektnyckel jämförs med objektnycklar som finns lagrade i organet för lagring för kontroll av överensstämmelse.

Objektnycklarna är företrädesvis sorterade efter antalet märkningar i organet för lagring.

- 5 Sökningen åstadkommes att resultera i ett antal på förhand begärda troliga kandidater för aktuellt sökt formulär.

En operatör kan vid behov manuellt stödja hela eller delar av den adaptiva registreringen eller identifieringen av nytt formulär resp registrerat formulär om flera alternativa kandidater förekommer som sannolika enligt ett godhetstal.

- 10 Slutligen kan formulärets identitet fastställas genom datafångst av ett RCG-värde. Vidare anger föreliggande uppfinning en anordning för utövande av ovanstående förfarande.

- 15 Anordningen genomför automatisk datafångst, via organ för detsamma, av formulär med på förhand okänd utformning och informationsinnehåll, vid inmatning i nämnda organ, samt lagring av förlagor av densamma. Den lär sig adaptivt och registrerar utformningen av formulär, och innefattar en dator med följande organ för att åstadkomma den adaptiva registreringen:

organ för alstring av en formulärkarta utifrån det på förhand okända formulärets utformning för identifiering av formulär innehållande information;

- 20 organ för sökning med jämförelse av formulärkartan med lagrade inlärd kartor i ett organ för lagring av formulärkartor;

organ för lagring av alstrad formulärkarta i organet för lagring då den ej överensstämmer med en lagrad karta enligt förutbestämda gränser för överensstämmelse;

- 25 organ för indikering av överensstämmelse enligt gränserna för överensstämmelse då överensstämmelse föreligger; och

organ för identifiering och fortsatt datafångst av formulärets informationsinnehåll.

Ytterligare kan anordningen innefatta eller utföra det som anges enligt förfarandet ovan i föreliggande uppfinning, vilket ytterligare belyses via bilagda osjälvständiga anordningskrav.

- 30 **Kortfattad beskrivning av ritningsfigurer**

Fortsättningsvis hänvisas till bilagda figurer med tillhörande text för en bättre förståelse av föreliggande uppfinning.

Fig. 1 illustrerar schematiskt hur ett linjemönster åstadkommes ur en inskannad faktura.

Fig. 2 illustrerar schematiskt ett flöde för skanning, identifiering, tolkning, validering av ett formulär enligt föreliggande uppfinning.

5 Detaljerad beskrivning av föredragna utföringsformer

I den fortsatta beskrivningen av föreliggande uppfinning förekommer formulär som fakturor. Uppfinningen är dock inte begränsad till fakturor som formulär, utan allehanda dokument innefattande text, figurer etc är att betrakta som formulär. Fakturor används här som ett exempel på ett formulär för exemplifiering av uppfinningen.

10 I fig. 1 illustreras schematiskt en del av en faktura 10 som skannats in i en dator och som visas på datorns display. Som framgår av fakturan 10 är den otydlig eller blurrig efter inskanningen eller avläsningen.

Fakturan 10 består till en del av en logotyp 12 och olika vertikala och horisontella linjeelement 14.

15 Att observera är även att logotypen innehåller långa svarta, eller färgade i skiftande kulörer, linjeelement 16 som delvis registreras i en linjekarta 18 enligt föreliggande uppfinning och ger en uppfattning av hur den ursprungliga logotypen 12 såg ut, vilket underlättar när fakturan är föremål för att identifieras i en formulärkartsdatabas som registrerad. I färgade linjeelement innefattas även gråskalor av svart.

20 Formulärkartan som här utgör en linjekarta 18 har filtrerats från andra objekt 19, som exempelvis hela eller delar av textobjekt eller färgobjekt, som nämnts även linjeelement innefattande färg, vilka här inte kan återges, men ett formulär 10 kan innefatta många färgfält.

25 En faktura 10 som bearbetas enligt föreliggande uppfinning, härafter benämnd EH (Eyes & Hands), måste i ett tidigt skede identifieras. För att en identifikation skall lyckas måste EH vid något tidigare tillfälle ha lärt sig hur den aktuella fakturan 10 ser ut, vilket i korthet innebär att informationen om fakturan finns i en formulärdatabas i EH.

30 Identifikationen måste med nödvändighet vara snabb och kunna åstadkommas mot en databas som innehåller ett mycket stort antal möjliga fakturaidentiteter 18. Det är inte ovanligt att det förekommer mer än 10 000 identiteter 18 i databasen.

Förfarandet och anordningen som EH använder kräver inte att en faktura alltid dras igenom en skanner på exakt samma sätt, d v s information på fakturan tillåts att

variera något i x- och y-led inom förutbestämda mått eller tröskelvärden. I fig. 1 visas schematiskt ett kartesiskt koordinatsystem.

I föreliggande uppfinning (EH) innefattas att EH i en utföringsform söker efter samtliga vertikala 14 och horisontella 15 linjeelement av en förutbestämd längd på

- 5 fakturan 10. Linjerna 14, 15 behöver inte vara fristående och isolerade, utan kan t ex vara en del i en större logotext 12, ReadSoft AB i fig. 1. Logotypen 12 avspeglar sig som linjeelement 16 i linjekartan 18.

De horisontella linjerna 15 och vertikala linjerna utgör grund för alstring av en horisontell nyckel (h-nyckel) resp en vertikal nyckel (v-nyckel) enligt följande:

10

* Fakturan indelas i ett stort antal horisontella segment längs y-axeln (ej visat). Varje segment motsvarar en position i h-nyckeln. Om ett visst segment innehåller ett eller flera linjeelement 15 sätts ett märke (tagg) på motsvarande nyckelposition, i annat fall en lucka, ett inverterat märke etc, i stort något som skiljer sig från märket.

15

* En v-nyckel för de vertikala linjeelementen 14 alstras på liknande sätt längs x-axeln.

* H- och v-nyckeln benämnes och utgör tillsammans en linjenyckel. Härfter åstadkommes en sökning, vilket innebär att aktuell linjenyckel jämförs med de linjenycklar, för kända fakturor 10, som finns i EH:s databas. Vid jämförelsen tas hänsyn till att enstaka linjer eller linjeelement 14, 15 kan variera något i position, eller i vissa fall helt saknas, samt att det totala linjemönstret kan vara förskjutet något enligt lämpliga förutbestämda värden i x- och y-led resp horisontellt och vertikalt.

20

* Linjenycklarna i databasen ligger sorterade efter antalet märken (taggar), vilket är till för att effektivisera sökningen.

* Sökningen i sig resulterar i, ett på förhand, antal troliga kandidater på aktuell fakturaidentitet 10. Samtliga kandidater är även förknippade med ett godhetstal eller en sannolikhet för att de är aktuell faktura 10.

25

* Fakturans identitet fastställs slutligen genom att åstadkomma en tolkning av det s k RCG-värdet (RCG-ReCoGnition). RCG-värdet är ett värde, på given position, som är unikt för en viss faktura/leverantör eller annat formulär. Exempel på

30

sådana värden är bankgironummer, postgironummer, fakturanummer, totalsumma etc.

Nämnda segment kan exempelvis bilda rutmönster mer eller mindre finmaskiga

5 efter behov relaterat till en snabb sökning.

Linjenycklar kan även implementeras på objekt i form av hel eller del av text och färger, vilka tilldelas linjenycklar ur en objektområdeslista som innefattar x- och y-nycklar för objekt. Objektområdeslistan kan exempelvis bestå av positioner för vissa utvalda objekt. Principerna för linjekartor enligt ovan kan även anammas för andra objekt än

10 linjeelement för att åstadkomma identifiering av formulär.

Vidare om linjenycklar inte återfinnes i databasen indikerar det att fakturan inte är känd, vilket får till följd att den nya linjenyckeln lagras i databasen som således uppdateras i realtid.

Vid behov kan en operatör manuellt via sin dator stödja hela eller del av adaptiv
15 registrering och/eller identifiering av nytt formulär resp registrerat formulär om flera alternativa kandidater förekommer som sannolika enligt godhetstalet.

Vidare innefattar föreliggande uppfinning en anordning för utövande av farandet enligt ovan.

Anordningen utför automatisk datafångst, via organ för detsamma, av formulär
20 med på förhand okänd utformning och informationsinnehåll, vid inmatning i nämnda organ, samt lagring av förlagor av densamma. Den registrerar adaptivt och lär sig utformningen av formulär, och innefattar en dator med följande organ för att åstadkomma den adaptiva registreringen:

organ för alstring av en formulärkarta utifrån det på förhand okända formulärets
25 utformning för identifiering av formulär innehållande information;

organ för sökning med jämförelse av formulärkartan med lagrade inlärd kartor i ett organ för lagring av formulärkartor;

organ för lagring av alstrad formulärkarta i organet för lagring då den ej överensstämmer med en lagrad karta enligt förutbestämda gränser för överensstämmelse;

30 organ för indikering av överensstämmelse enligt gränserna för överensstämmelse då överensstämmelse föreligger; och

organ för identifiering och fortsatt datafångst av formulärets informationsinnehåll.

Nämnda organ är företrädesvis hård- och/eller programvarustyrda, såsom exempelvis:

Skanner för datafångst.

Elektroniska lagringsmedium (hårddisk, CD-skiva etc) för organet för lagring.

5 Tecken, ikoner, signalgeneratorer etc som organ för indikering.

Filter och jämförare för organ för sökning och jämförelse samt filter och register för identifiering.

Överhuvudtaget är organen som utnyttjas i föreliggande uppfinning kända för en fackman inom teknikområdet, men deras samordning för att uppnå uppfinningens syften
10 och ändamål är däremot innovativ.

I en utföringsform av föreliggande uppfinning med hänvisning till fig. 2 illustreras schematiskt ett flödesschema för skanning, identifiering, tolkning, validering av ett formulär enligt föreliggande uppfinning.

Fig. 2 är indelad i prickade delområden för att tydliggöra de olika stegen i ett
15 förfarande enligt föreliggande uppfinning, varvid stegen utgörs av skanning av formulär 200, identifiering av formulär 210, tolkning av formulär 220 samt validering av formulär 230.

Formuläret skannas 200 in i EH, varefter identifiering 210 följer. Identifieringen består av alstring av en linjekarta 212, alternativt en objektområdeslista, varvid en
20 linjenyckel alstras. Därefter jämförs 214 formulär 10 mot kända nycklar i formulärkarts-databasen, varvid en bekräftelse av identifieringen erhålles via RCG-värdet. Nästa steg innefattar att avgöra om identifieringen lyckades 216 enligt villkoret "Ja" eller "Nej". Om avgörandet resulterar i "Nej" undersöks villkorligt om det finns fler kandidater i form av linjenycklar 218. Om svaret är "Ja" genomlöpes en slinga i form av 214, 216, 218 ända
25 tills lyckad identifiering åstadkommes 216 eller att det inte förekommer fler linjenyckel-kandidater 218.

I det fall att identifieringen lyckas vidtar tolkningen 220 av formuläret genom tolkning med hjälp av aktuell formulärkarta 222, varefter validering 230 eller utvärdering av formulärs 10 fält sker 232. Som en option kan en operatör hjälpa till med urvalet om
30 flera alternativa fält återfinnes 234.

I fall att identifiering 210 misslyckas och att det inte förekommer fler linjenyckel-kandidater 218 utförs tolkning 220 genom att självupplärning med en formulärdefinition

224 åstadkommes. Formulärdefinitionen består av en mall eller ett regelverk som beskriver det som är gemensamt för en specifik samling av formulär, exempelvis svenska fakturor. Därefter tolkas RCG-värdet 226 och det avgörs 228 om aktuellt RCG-värde återfinnes i formulärkartsdatabasen. Om svaret är "Ja" vidtar en omtolkning 229, varefter
5 tolkningen fortsätter vid 222 för att leda till valideringen 232.

Om däremot svaret är "Nej" vidtar validering 230, 236, varefter formuläret sparas i formulärkartsdatabasen med linjenyckeln 238. Innan stegen 236, 238 genomföres kan även en operatör åstadkomma upplärning 239, om flera alternativa fält återfinnes.

Ovan beskrivna utföringsformer av föreliggande uppfinning är möjliga sådana,
10 men inte ämnade att begränsa uppfinningen som sådan, utan för fackmannen inom teknikområdet framgår ytterligare utföringsformer via avfattningen av bilagda patentkrav.

Patentkrav

1. Förfarande för automatisk datafångst (200), via organ för detsamma, av formulär (10) med på förhand okänd utformning och informationsinnehåll (19), vid inmatning i nämnda organ, samt lagring av förlagor av densamma, k ä n n e t e c k n a t
5 av att förfarandet är adaptivt (224) och innefattar inläring och registrering av utformningen av formulär, varvid det innefattar följande steg för att åstadkomma adaptiv registrering (238):

alstring av en formulärkarta (18) utifrån det på förhand okända formulärets (10) utformning för identifiering (210) av formulär innehållande information;

10 sökning med jämförelse (210, 220) av formulärkartan (18) med lagrade inlärd kartor i ett organ för lagring av formulärkartor;

lagring (238) av alstrad formulärkarta (18) i organet för lagring då den ej överensstämmer med en lagrad karta enligt förutbestämda gränser för överensstämmelse;

15 indikering av överensstämmelse enligt gränserna för överensstämmelse då överensstämmelse föreligger; och

fortsatt datafångst (232) för identifiering av formulärets informationsinnehåll.

2. Förfarande enligt krav 1, k ä n n e t e c k n a t av att formulärkartan (18) består av en objektområdeslista med objekt (19) innefattade i formulär.

20 3. Förfarande enligt krav 2, k ä n n e t e c k n a t av att objekten (19) består av färger och/eller hel eller del av text.

4. Förfarande enligt krav 1, k ä n n e t e c k n a t av att formulärkartan (18) utgör en linjekarta innefattande linjeelement (14, 15) ur formuläret (10).

25 5. Förfarande enligt krav 4, k ä n n e t e c k n a t av att horisontella linjer (15) i linjekartan används för alstring av en horisontell nyckel genom att formuläret indelas i ett förutbestämt antal horisontella segment längs y-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den horisontella nyckeln.

30 6. Förfarande enligt krav 4 och 5, k ä n n e t e c k n a t av att vertikala linjer (14) i linjekartan används för alstring av en vertikal nyckel genom att formuläret (10) indelas i ett förutbestämt antal vertikala segment längs x-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den vertikala nyckeln.

7. Förfarande enligt krav 5 och 6, k ä n n e t e c k n a t av att minst ett linjeelement (14, 15) som innefattas i ett segment märks i motsvarande nyckelposition och

att segment vilka saknar linjeelement förblir omärkta i motsvarande nyckelposition.

8. Förfarande enligt krav 4-7, k ä n n e t e c k n a t av att en horisontell nyckel och/eller en vertikal nyckel utgör en linjenyckel i linjekartan (18), varvid, vid nämnda sökning, alstrad linjenyckel jämförs (214) med linjenycklar som finns lagrade i organet för lagring för kontroll av överensstämmelse.

9. Förfarande enligt krav 8, k ä n n e t e c k n a t av att linjenycklarna är sorterade efter antalet märkningar i organet för lagring.

10. Förfarande enligt krav 1 och 2, k ä n n e t e c k n a t av att objekts horisontella position i objektområdeslistan används för alstring av en horisontell nyckel genom att formuläret indelas i ett förutbestämt antal horisontella segment längs y-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den horisontella nyckeln.

11. Förfarande enligt krav 1 och 2, k ä n n e t e c k n a t av att objekts vertikala position i objektområdeslistan används för alstring av en vertikal nyckel genom att formuläret indelas i ett förutbestämt antal vertikala segment längs x-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den vertikala nyckeln.

12. Förfarande enligt krav 10 och 11, k ä n n e t e c k n a t av att minst ett objekt som innefattas i ett segment märks i motsvarande nyckelposition och att segment vilka saknar objekt förblir omärkta i motsvarande nyckelposition.

13. Förfarande enligt krav 10-12, k ä n n e t e c k n a t av att en horisontell nyckel och/eller en vertikal nyckel utgör en objektnyckel i objektområdeslistan, varvid, vid nämnda sökning, alstrad objektnyckel jämförs med objektnycklar som finns lagrade i organet för lagring för kontroll av överensstämmelse.

14. Förfarande enligt krav 15, k ä n n e t e c k n a t av att objektnycklarna är sorterade efter antalet märkningar i organet för lagring.

15. Förfarande enligt krav 1-14, k ä n n e t e c k n a t av att sökningen (210, 218) åstadkommes att resultera i ett antal på förhand begärda troliga kandidater för aktuellt sökt formulär.

16. Förfarande enligt krav 15, k ä n n e t e c k n a t av att en operatör manuellt (234, 239) stöder hela eller delar av den adaptiva registreringen eller identifieringen av nytt formulär resp registrerat formulär om flera alternativa kandidater före-

kommer som sannolika enligt ett godhetstal.

17. Förfarande enligt krav 1-16, k ä n n e t e c k n a t av att formulärets identitet fastställs genom datafångst av RCG-värdet (214).

18. Anordning för automatisk datafångst, via organ för detsamma, av formulär
5 (10) med på förhand okänd utformning och informationsinnehåll (19), vid inmatning i nämnda organ, samt lagring av förlagor av densamma, k ä n n e t e c k n a d av att den adaptivt lär sig och registrerar utformningen av formulär (10), och innefattar en dator med följande organ för att åstadkomma den adaptiva registreringen (238):

organ för alstring av en formulärkarta (18) utifrån det på förhand okända formulä-
10 rets (10) utformning för identifiering av formulär innehållande information;

organ för sökning med jämförelse av formulärkartan med lagrade inlärd kartor i ett organ för lagring av formulärkartor;

organ för lagring av alstrad formulärkarta (18) i organet för lagring då den ej överensstämmer med en lagrad karta enligt förutbestämda gränser för överensstämmelse;

15 organ för indikering av överensstämmelse enligt gränserna för överensstämmelse då överensstämmelse föreligger; och

organ för identifiering och fortsatt datafångst av formulärets informationsinnehåll.

19. Anordning enligt krav 18, k ä n n e t e c k n a d av att formulärkartan (18) består av en objektområdeslista med objekt (19) innefattade i formulär (10).

20 20. Anordning enligt krav 19, k ä n n e t e c k n a d av att objekten (19) består av färger och/eller hel eller del av text.

21. Anordning enligt krav 18, k ä n n e t e c k n a d av att formulärkartan (18) utgör en linjekarta innefattande linjeelement (14, 15) ur formuläret.

22. Anordning enligt krav 21, k ä n n e t e c k n a d av att horisontella linjer
25 (15) i linjekartan används för alstring av en horisontell nyckel genom att formuläret indelas i ett förutbestämt antal horisontella segment längs y-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den horisontella nyckeln.

23. Anordning enligt krav 21 och 22, k ä n n e t e c k n a d av att vertikala linjer (14) i linjekartan används för alstring av en vertikal nyckel genom att formuläret
30 indelas i ett förutbestämt antal vertikala segment längs x-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den vertikala nyckeln.

24. Anordning enligt krav 22 och 23, k ä n n e t e c k n a d av att minst ett

linjeelement (14, 15) som innefattas i ett segment märks i motsvarande nyckelposition och att segment vilka saknar linjeelement förblir omärkta i motsvarande nyckelposition.

25. Anordning enligt krav 22-24, k ä n n e t e c k n a d av att en horisontell nyckel och/eller en vertikal nyckel utgör en linjenyckel i linjekartan, varvid, vid nämnda
5 sökning, alstrad linjenyckel jämförs med linjenycklar som finns lagrade i organet för lagring för kontroll av överensstämmelse.

26. Anordning enligt krav 25, k ä n n e t e c k n a d av att linjenycklarna är sorterade efter antalet märkningar i organet för lagring.

27. Anordning enligt krav 18 och 19, k ä n n e t e c k n a d av att objekts
10 (19) horisontella position i objektområdeslistan används för alstring av en horisontell nyckel genom att formuläret indelas i ett förutbestämt antal horisontella segment längs y-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den horisontella nyckeln.

28. Anordning enligt krav 18 och 19, k ä n n e t e c k n a d av att objekts
15 (19) vertikala position i objektområdeslistan används för alstring av en vertikal nyckel genom att formuläret indelas i ett förutbestämt antal vertikala segment längs x-axeln i ett kartesiskt koordinatsystem, varvid varje segment motsvarar en position i den vertikala nyckeln.

29. Anordning enligt krav 27 och 28, k ä n n e t e c k n a d av att minst ett
20 objekt som innefattas i ett segment märks i motsvarande nyckelposition och att segment vilka saknar objekt förblir omärkta i motsvarande nyckelposition.

30. Anordning enligt krav 27-29, k ä n n e t e c k n a d av att en horisontell nyckel och/eller en vertikal nyckel utgör en objektnyckel i objektområdeslistan, varvid, vid nämnda sökning, alstrad objektnyckel jämförs med objektnycklar som finns lagrade i
25 organet för lagring för kontroll av överensstämmelse.

31. Anordning enligt krav 30, k ä n n e t e c k n a d av att objektnycklarna är sorterade efter antalet märkningar i organet för lagring.

32. Anordning enligt krav 18-31, k ä n n e t e c k n a d av att sökningen åstadkommes att resultera i ett antal på förhand begärda troliga kandidater för aktuellt sökt
30 formulär.

33. Anordning enligt krav 15, k ä n n e t e c k n a d av att en operatör via anordningen manuellt kan stödja hela eller delar av den adaptiva registreringen eller

identifieringen av nytt formulär resp registrerat formulär om flera alternativa kandidater förekommer som sannolika enligt ett godhetstal.

34. Anordning enligt krav 18-33, k ä n n e t e c k n a t av att formulärets identitet fastställs genom datafångst av RCG-värdet.

5 -----

Sammandrag

Uppfinningen anger ett förfarande och en anordning för automatisk datafångst av pappersburen information, varvid ett okänt formulär (10) som skannas in i en dator omvandlas till linjeelement (14, 15) för snabbare identifiering. Identifiering av kända
5 blanketter (10) underlättas även via linjeelementen (14, 15).

(Fig. 1)

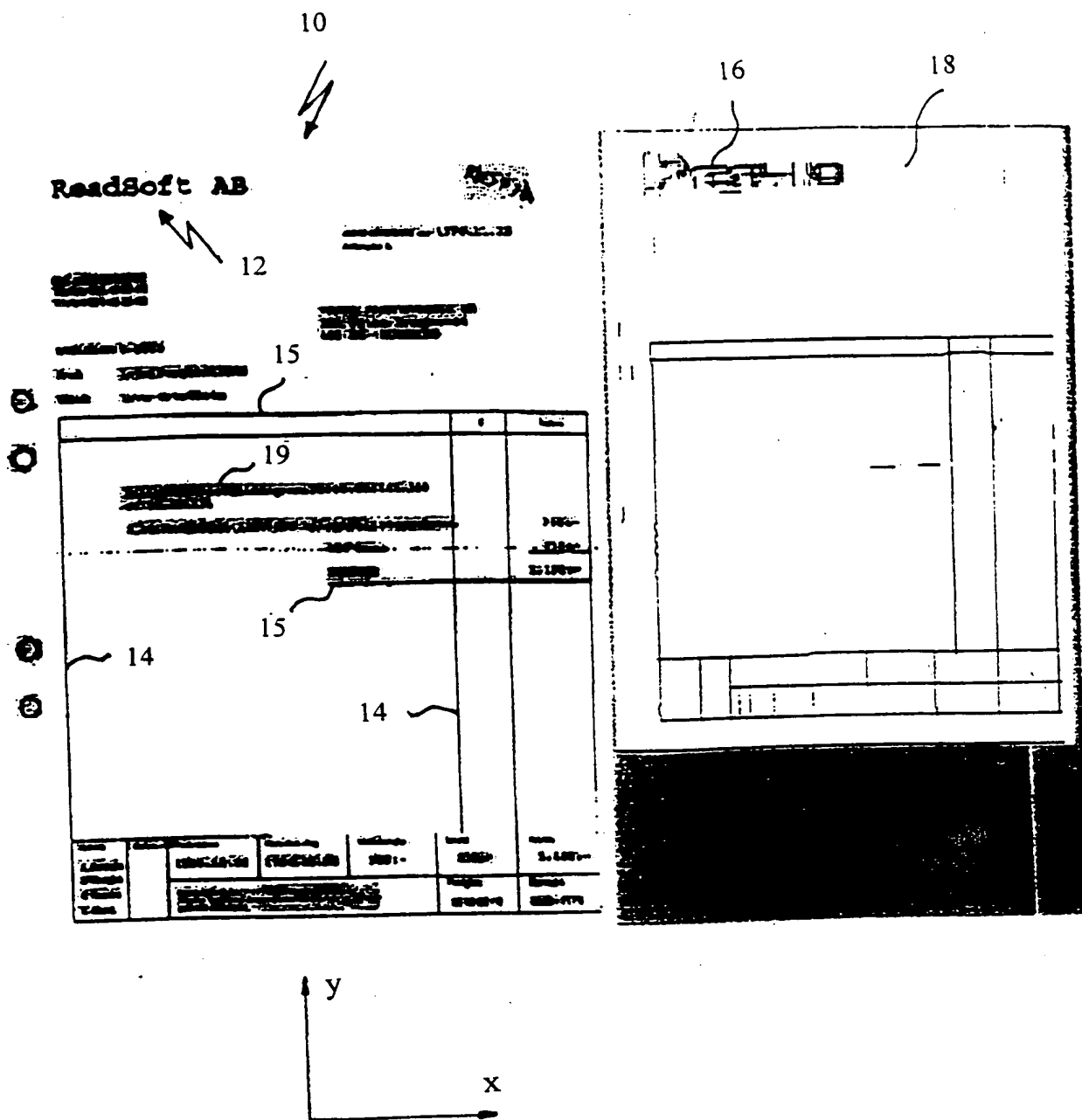


Fig. 1

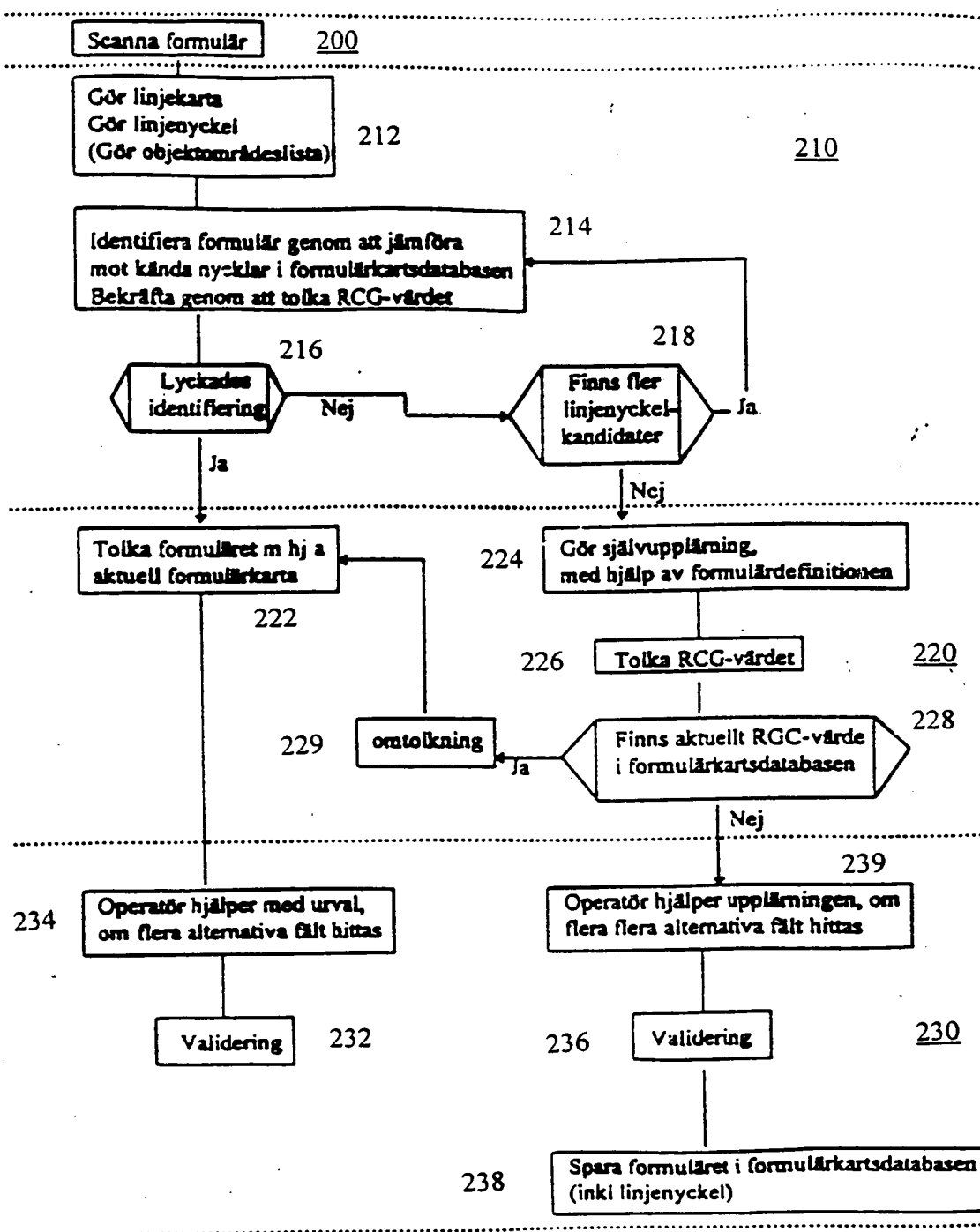


Fig. 2

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No.

PCT/SE98/00602

International Filing Date

April 1, 1998

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference
(if desired) (12 characters maximum)

P33354PC00

Box No. I TITLE OF INVENTION

Method and arrangement for automatic data acquisition of forms

Box No. II APPLICANT

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)

ReadSoft AB
Garnisionsgatan 25A
S-254 66 HELSINGBORG
Sweden

☐ This person is also inventor.

Telephone No.

Facsimile No.

Teleprinter No.

State (i.e. country) of nationality:

SE

State (i.e. country) of residence:

SE

This person is applicant
for the purposes of:

☐

all designated
States

☒

all designated States except
the United States of America

☐

the United States
of America only

☐

the States indicated in
the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)

ANDERSSON, Jan
Banckagatan 17
S-260 40 VIKEN
Sweden

This person is:

☐ applicant only

☒ applicant and inventor

☐ inventor only (If this check-box
is marked, do not fill in below.)

State (i.e. country) of nationality:

SE

State (i.e. country) of residence:

SE

This person is applicant
for the purposes of:

☐

all designated
States

☐

all designated States except
the United States of America

☒

the United States
of America only

☐

the States indicated in
the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:

☒

agent

☐

common representative

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)

AB STOCKHOLMS PATENTBYRÅ, Zacco & Bruhn (publ)
ONN, Thorsten; AGVALD-GLAS, Gunilla; BERNHULT, Lennart;
HANSSON, Sven; PETRÉ, Urban; WESTERLUND, Örjan
Box 23101, S-104 35 STOCKHOLM, Sweden

Telephone No.

08-7299500

Facsimile No.

08-318315

Teleprinter No.

☐ Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Box No.V DESIGNATION OF STATE

The following designations are hereby made under Rule 4.9(a) (mark the applicable check-boxes; at least one must be marked):

Regional Patent

- ☒ AP ARIPO Patent: GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SZ Swaziland, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☒ EA Eurasian Patent: AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☒ EP European Patent: AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☒ OA OAPI Patent: BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT (if other kind of protection or treatment desired, specify on dotted line)

National Patent (if other kind of protection or treatment desired, specify on dotted line):

- | | |
|--|--|
| <input checked="" type="checkbox"/> AL Albania | <input checked="" type="checkbox"/> LT Lithuania |
| <input checked="" type="checkbox"/> AM Armenia | <input checked="" type="checkbox"/> LU Luxembourg |
| <input checked="" type="checkbox"/> AT Austria | <input checked="" type="checkbox"/> LV Latvia |
| <input checked="" type="checkbox"/> AU Australia | <input checked="" type="checkbox"/> MD Republic of Moldova |
| <input checked="" type="checkbox"/> AZ Azerbaijan | <input checked="" type="checkbox"/> MG Madagascar |
| <input checked="" type="checkbox"/> BA Bosnia and Herzegovina | <input checked="" type="checkbox"/> MK The former Yugoslav Republic of Macedonia |
| <input checked="" type="checkbox"/> BB Barbados | |
| <input checked="" type="checkbox"/> BG Bulgaria | <input checked="" type="checkbox"/> MN Mongolia |
| <input checked="" type="checkbox"/> BR Brazil | <input checked="" type="checkbox"/> MW Malawi |
| <input checked="" type="checkbox"/> BY Belarus | <input checked="" type="checkbox"/> MX Mexico |
| <input checked="" type="checkbox"/> CA Canada | <input checked="" type="checkbox"/> NO Norway |
| <input checked="" type="checkbox"/> CH and LI Switzerland and Liechtenstein | <input checked="" type="checkbox"/> NZ New Zealand |
| <input checked="" type="checkbox"/> CN China | <input checked="" type="checkbox"/> PL Poland |
| <input checked="" type="checkbox"/> CU Cuba | <input checked="" type="checkbox"/> PT Portugal |
| <input checked="" type="checkbox"/> CZ Czech Republic | <input checked="" type="checkbox"/> RO Romania |
| <input checked="" type="checkbox"/> DE Germany | <input checked="" type="checkbox"/> RU Russian Federation |
| <input checked="" type="checkbox"/> DK Denmark | <input checked="" type="checkbox"/> SD Sudan |
| <input checked="" type="checkbox"/> EE Estonia | <input checked="" type="checkbox"/> SE Sweden |
| <input checked="" type="checkbox"/> ES Spain | <input checked="" type="checkbox"/> SG Singapore |
| <input checked="" type="checkbox"/> FI Finland | <input checked="" type="checkbox"/> SI Slovenia |
| <input checked="" type="checkbox"/> GB United Kingdom | <input checked="" type="checkbox"/> SK Slovakia |
| <input checked="" type="checkbox"/> GE Georgia | <input checked="" type="checkbox"/> SL Sierra Leone |
| <input checked="" type="checkbox"/> GH Ghana | <input checked="" type="checkbox"/> TJ Tajikistan |
| <input checked="" type="checkbox"/> GM Gambia | <input checked="" type="checkbox"/> TM Turkmenistan |
| <input checked="" type="checkbox"/> GW Guinea-Bissau | <input checked="" type="checkbox"/> TR Turkey |
| <input checked="" type="checkbox"/> HU Hungary | <input checked="" type="checkbox"/> TT Trinidad and Tobago |
| <input checked="" type="checkbox"/> ID Indonesia | <input checked="" type="checkbox"/> UA Ukraine |
| <input checked="" type="checkbox"/> IL Israel | <input checked="" type="checkbox"/> UG Uganda |
| <input checked="" type="checkbox"/> IS Iceland | <input checked="" type="checkbox"/> US United States of America |
| <input checked="" type="checkbox"/> JP Japan | |
| <input checked="" type="checkbox"/> KE Kenya | <input checked="" type="checkbox"/> UZ Uzbekistan |
| <input checked="" type="checkbox"/> KG Kyrgyzstan | <input checked="" type="checkbox"/> VN Viet Nam |
| <input checked="" type="checkbox"/> KP Democratic People's Republic of Korea | <input checked="" type="checkbox"/> YU Yugoslavia |
| | <input checked="" type="checkbox"/> ZW Zimbabwe |
| <input checked="" type="checkbox"/> KR Republic of Korea | |
| <input checked="" type="checkbox"/> KZ Kazakhstan | |
| <input checked="" type="checkbox"/> LC Saint Lucia | |
| <input checked="" type="checkbox"/> LK Sri Lanka | |
| <input checked="" type="checkbox"/> LR Liberia | |
| <input checked="" type="checkbox"/> LS Lesotho | |

Check-boxes reserved for designating States (for the purposes of a national patent) which have become party to the PCT after issuance of this sheet:

- ☐
- ☐
- ☐

In addition to the designations made above, the applicant also makes under Rule 4.9(b) all designations which would be permitted under the PCT except the designation(s) of

The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. (Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SE 98/00602

A. CLASSIFICATION OF SUBJECT MATTER		
IPC6: G06K 9/20 // G 06 F 17/24 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
IPC6: G06F, G06K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
SE,DK,FI,NO classes as above		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
WPI		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4933979 A (A. SUZUKI ET AL.), 12 June 1990 (12.06.90), column 2, line 10 - line 51; column 9, line 51 - column 10, line 52 --	1-34
X	US 5140650 A (R.G. CASEY ET AL.), 18 August 1992 (18.08.92), column 2, line 18 - line 65 --	1-34
A	US 5293429 A (A. PIZANO ET AL.), 8 March 1994 (08.03.94), abstract -----	1-34
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
2 July 1998		06-07-1998
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86		Authorized officer Malin Keijser Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

Information on patent family members

09/06/98

International application No.

PCT/SE 98/00602

Patent document cited in search report			Publication date	Patent family member(s)		Publication date	
US	4933979	A	12/06/90	JP	63155386 A	28/06/88	

US	5140650	A	18/08/92	CA	2000012 A,C	02/08/90	
				JP	2231692 A	13/09/90	

US	5293429	A	08/03/94	JP	5217019 A	27/08/93	

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

United States Patent and Trademark
Office
(Box PCT)
Crystal Plaza 2
Washington, DC 20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 24 November 1998 (24.11.98)	
International application No. PCT/SE98/00602	Applicant's or agent's file reference P33354PC00
International filing date (day/month/year) 01 April 1998 (01.04.98)	Priority date (day/month/year) 01 April 1997 (01.04.97)
Applicant ANDERSSON, Jan	

1. The designated Office is hereby notified of its election made:



in the demand filed with the International Preliminary Examining Authority on:

29 October 1998 (29.10.98)



in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was



was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

<p>The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland</p> <p>Facsimile No.: (41-22) 740.14.35</p>	<p>Authorized officer E. Huchon</p> <p>Telephone No.: (41-22) 338.83.38</p>
--	---

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference P33354PC00	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/SE98/00602	International filing date (<i>day/month/year</i>) 01.04.1998	Priority date (<i>day/month/year</i>) 01.04.1997
International Patent Classification (IPC) or national classification and IPC ₆ G06K 9/20 // G06F 17/24		
Applicant ReadSoft AB et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 3 sheets, including this cover sheet.

☒ This report is also accompanied by ANNEXES; i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 5 sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand 29.10.1998	Date of completion of this report 16.04.1999
Name and mailing address of the IPEA/SE Patent- och registreringsverket Box 5055 S-102 42 STOCKHOLM Facsimile No. 08-667 72 88	Authorized officer Malin Keijser Telephone No. 08-782 25 00

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/SE98/00602

I. Basis of the report

1. This report has been drawn on the basis of *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.)*:

☐ the international application as originally filed.

☒ the description, pages 1-9, as originally filed,
 pages _____, filed with the demand,
 pages _____, filed with the letter of _____,
 pages _____, filed with the letter of _____.

☒ the claims, Nos. _____, as originally filed,
 Nos. _____, as amended under Article 19,
 Nos. _____, filed with the demand,
 Nos. 1-34, filed with the letter of 14.04.1999,
 Nos. _____, filed with the letter of _____.

☒ the drawings, sheets/fig 1-2, as originally filed,
 sheets/fig _____, filed with the demand
 sheets/fig _____, filed with the letter of _____,
 sheets/fig _____, filed with the letter of _____.

2. The amendments have resulted in the cancellation of:

☐ the description, pages _____

☐ the claims, Nos. _____

☐ the drawings, sheets/fig _____

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the supplemental Box (Rule 70.2(c)).

4. Additional observations, if necessary:

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/SE98/00602

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**1. Statement**

Novelty (N)	Claims	<u>1-34</u>	YES
	Claims	_____	NO
Inventive step (IS)	Claims	<u>1-34</u>	YES
	Claims	_____	NO
Industrial applicability (IA)	Claims	<u>1-34</u>	YES
	Claims	_____	NO

2. Citations and explanations

The claimed invention relates to a method and a device for automatic data acquisition of forms where the system has no prior knowledge about what the form looks like.

In the International Search Report, the following documents were cited:

D1: US, A, 4933979

D2: US, A, 5140650

D3: US, A, 5293429

D1 describes a data reading apparatus which registers form information for each kind of form sheet from which the data are to be read (see column 9, line 51 - column 10, line 52).

D2 shows a method for automatic extraction of data from printed forms. A blank master form is first scanned and stored.

In D3, a system and a method for automatically classifying forms is described.

The described invention differs from prior art according to the cited documents in that no initial scanning of a blank form sheet is necessary. This difference cannot be considered as obvious to a person skilled in the art. The subject matter of claims 1-34 is therefore novel and inventive.

Consequently, the subject matter of claims 1-34 is considered to fulfil the requirements of novelty, inventive step and industrial applicability.

1. Method for the automatic data acquisition (200), by means of a means for the same, of forms (10) whose design and information content (19) is not known in advance, by input into the said means, together with storage of patterns of the same, characterised in that the method is adaptive (224) and includes self-learning and registration of the design of forms without initial scanning of a blank form sheet being necessary, whereby it includes the following steps to accomplish the adaptive registration (238):

generation of a form map (18) based on the previously unknown form's (10) design for identifying (210) information contained on the form;

searching and comparing (210, 220) the form map (18) with stored, registered maps in a means for storing form maps;

storage (238) of generated form maps (18) in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

indication of agreement according to the limits for agreement when agreement is found; and

continued data acquisition (232) for identifying the information content of the form.

2. Method according to claim 1, characterised in that the form map (18) consists of an object area list with objects (19) contained in the form.

3. Method according to claim 2, characterised in that the object (19) comprises colours and/or wholly or partly of text.

4. Method according to claim 1, characterised in that the form map (18) constitutes a line map comprising line elements (14,15) from the form (10).

5. Method according to claim 4, characterised in that the horizontal lines (15) in the line map are used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of coordinates, whereby each segment is equivalent to a position in the horizontal key.

6. Method according to claims 4 and 5, characterised in that the vertical lines (14) in the line map are used to generate a vertical key by dividing the form (10) into a pre-determined number of vertical segments along the x-axis in a cartographic system of coordinates, whereby each segment is equivalent to a position in the vertical key.

7. Method according to claims 5 and 6, characterised in that at least one line element (14, 15) that is included in a segment is marked in the equivalent key position, and that segments that lack line elements remain unmarked in the equivalent key position.

5 8. Method according to claims 4-7, characterised in that the horizontal key and/or a vertical key constitute a line key in the line map (18), whereby, during the said searching, the line key generated is compared (214) with line keys stored in the storage means for verifying agreement.

9. Method according to claim 8, characterised in that the line keys are sorted in the storage means according to the number of markings.

10 10. Method according to claims 1 and 2, characterised in that the object's horizontal position in the object area list is used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

15 11. Method according to claims 1 and 2, characterised in that the object's vertical position in the object area list is used to generate a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

20 12. Method according to claims 10 and 11, characterised in that at least one object that is included in a segment is marked in the equivalent key position, and that segments that lack objects remain unmarked in the equivalent key position

13. Method according to claims 10-12, characterised in that a horizontal key and/or a vertical key constitute an object key in the object area list, whereby, during the said searching, the object key generated is compared with object keys stored in the storage means for verifying agreement.

25 14. Method according to claim 13, characterised in that the object keys are sorted in the storage means according to the number of markings.

15. Method according to claims 1-14, characterised in that the searching (210, 218) results in a pre-defined number of requested probable candidates for the currently searched form.

30 16. Method according to claim 15, characterised in that an operator can support manually (234, 239) the whole or parts of the adaptive registration or identification of the

AMENDED SHEET

14-04-1999

new form or registered forms respectively if several alternative candidates are found as probabilities according to a factor of merit.

17. Method according to claims 1-16, characterised in that the identity of the form is confirmed by the data acquisition of a ReCoGnition (RCG) value (214) which
5 uniquely identifies a form.

18. Arrangement for the automatic data acquisition, by means of a means for the same, of forms (10) whose design and information content (19) is not known in advance, by input into the said means together with storage of patterns of the same, characterised in that it learns adaptively and registers the design of forms (10) without initial scanning of a blank form
10 sheet being necessary, and includes a computer with the following means for carrying out the adaptive registration (238):

means for generating a form map (18) based on the previously unknown form's (10) design for identifying information contained on the form;

means for searching and comparing the form map with stored, recognised maps in a
15 means for storing form maps;

means for storage of generated form maps (18) in the storage means when they do not coincide with a stored map according to pre-determined limits for agreement;

means for indicating agreement according to the limits for agreement when agreement is found; and
20

means for identification and continued data acquisition of the information content of the form.

19. Arrangement according to claim 18 characterised in that the form map (18) consists of an object area list with objects (19) contained in the form (10).

20. Arrangement according to claim 19, characterised in that the object
25 (19) comprises colours and/or wholly or partly of text.

21. Arrangement according to claim 18, characterised in that the form map (18) constitutes a line map comprising line elements (14, 15) from the form.

22. Arrangement according to claim 21, characterised in that the horizontal lines (15) in the line map are used to generate a horizontal key by dividing the form
30 into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

AMENDED SHEET

23. Arrangement according to claims 21 and 22, characterised in that the vertical lines (14) in the line map are used to generate a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

5 24. Arrangement according to claims 22 and 23, characterised in that at least one line element (14, 15) that is included in a segment is marked in the equivalent key position, and that segments that lack line elements remain unmarked in the equivalent key position.

10 25. Arrangement according to claims 22-24, characterised in that the horizontal key and/or a vertical key constitute a line key in the line map, whereby, during the said searching, the line key generated is compared with line keys stored in the storage means for verifying agreement.

26. Arrangement according to claim 25, characterised in that the line keys are sorted in the storage means according to the number of markings.

15 27. Arrangement according to claims 18 and 19, characterised in that the object's (19) horizontal position in the object area list is used to generate a horizontal key by dividing the form into a pre-determined number of horizontal segments along the y-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the horizontal key.

20 28. Arrangement according to claims 18 and 19, characterised in that the object's (19) vertical position in the object area list is used to generate a vertical key by dividing the form into a pre-determined number of vertical segments along the x-axis in a cartographic system of co-ordinates, whereby each segment is equivalent to a position in the vertical key.

25 29. Arrangement according to claims 27 and 28, characterised in that at least one object that is included in a segment is marked in the equivalent key position, and that segments that lack objects remain unmarked in the equivalent key position.

30 30. Arrangement according to claims 27-29, characterised in that a horizontal key and/or a vertical key constitute an object key in the object area list, whereby, during the said searching, the object key generated is compared with object keys stored in the storage means for verifying agreement.

31. Arrangement according to claim 30, characterised in that the object keys are sorted in the storage means according to the number of markings.

14-04-1999

32. Arrangement according to claims 18-31, characterised in that the searching results in a pre-defined number of requested probable candidates for the currently searched form.

5 33. Arrangement according to claim 32, characterised in that an operator can support manually the whole or parts of the adaptive registration or identification of the new form or registered forms respectively if several alternative candidates are found as probabilities according to a factor of merit.

10 34. Arrangement according to claims 18-33, characterised in that the identity of the form is confirmed by the data acquisition of a a ReCoGnition (RCG) value (214) which uniquely identifies a form.

AMENDED SHEET